



Gregoire, L.J., Valdes, P.J., Payne, A.J., & Kahana, R. (2011). Optimal tuning of a GCM using modern and glacial constraints. *Climate Dynamics*, 37(3-4), 705 - 719. <https://doi.org/10.1007/s00382-010-0934-8>

Publisher's PDF, also known as Version of record

Link to published version (if available):
[10.1007/s00382-010-0934-8](https://doi.org/10.1007/s00382-010-0934-8)

[Link to publication record in Explore Bristol Research](#)
PDF-document

Copyright: The Author(s) 2010. This article is published with open access at Springerlink.com

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Optimal tuning of a GCM using modern and glacial constraints

Lauren J. Gregoire · Paul J. Valdes ·
Antony J. Payne · Ron Kahana

Received: 18 January 2010 / Accepted: 14 October 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract In climate models, many parameters used to resolve subgrid scale processes can be adjusted through a tuning exercise to fit the model's output to target climatologies. We present an objective tuning of a low resolution Atmosphere–Ocean General Circulation Model (GCM) called FAMOUS where ten model parameters are varied together using a Latin hypercube sampling method to create an ensemble of 100 models. The target of the tuning consists of a wide range of modern climate diagnostics and also includes glacial tropical sea surface temperature. The ensemble of models created is compared to the target using an Arcsin Mielke score. We investigate how the tuning method used and the addition of glacial constraints impact on the present day and glacial climates of the chosen models. Rather than selecting a single configuration which optimises the metric in all the diagnostics, we obtain a subset of nine ‘good’ models which display great differences in their climate but which, in some sense, are all better than the original configuration. In those simulations, the global temperature response to last glacial maximum forcings is enhanced compared to the control simulation and the glacial Atlantic Ocean circulation is more in agreement with observations. Our study demonstrates that selecting a single ‘optimal’ configuration, relying only on present day constraints may lead to misrepresenting climates different to that of today.

Keywords GCM · Tuning · Latin hypercube sampling · Last glacial maximum · Sea surface temperatures · Palaeo-proxies

1 Introduction

In climate models, many parameterisations used to resolve sub-grid scale processes use parameters poorly constrained by observations or that can depend on the resolution of the model. Model calibration, often called tuning, is a part of the model development process which consist of searching for the optimal parameter values that will minimise a metric (or cost function) representing the discrepancy between observations and model output. Because General Circulation Models (GCMs) are computationally demanding, these models are generally ‘hand tuned’ often varying one parameter at a time, having a strong limitation on the number of tests that can be run and therefore relying heavily on expert knowledge. More systematic methods of parameter estimation that necessitate a large number of simulations have been developed and used on intermediate complexity models (Annan and Hargreaves 2007). With the increase of computing resources, systematic methods of tuning can now be applied to low resolution GCMs such as FAMOUS.

Jones et al. (2005) performed a systematic tuning of FAMOUS, using an iterative algorithm. Smith et al. (2008) further tuned the model manually by changing other parameters in the model. FAMOUS was intended to be a fast version of HadCM3 and was therefore tuned towards equivalent HadCM3 results in both studies. However, FAMOUS is increasingly being used for palaeoclimate studies where it can be argued that a better tuning target is present day observations and palaeo-data. Our aim was to tune the version of FAMOUS used for modelling quaternary climate towards observational present day and Last Glacial Maximum (LGM; 21 kyr B.P.) data.

The methodology used by Jones et al. (2005) is a successive minimisation algorithm. Such methodology is inadequate when parameters are correlated with each other

L. J. Gregoire (✉) · P. J. Valdes · A. J. Payne · R. Kahana
School of Geographical Sciences, University of Bristol,
Bristol, UK
e-mail: Lauren.gregoire@bristol.ac.uk

which is the case for cloud parameters in FAMOUS. Three other types of parameter estimation methods can be used in climate modelling (Annan and Hargreaves 2007). The simplest methods consist of sampling the whole parameter space. This method usually requires a number of samples which increases exponentially with the number of parameters but the efficiency of the sampling can be improved by using a Latin Hypercube Sampling (LHS; McKay 1992) which has been successfully used in uncertainty analysis studies (Schneider von Deimling et al. 2006; Edwards and Marsh 2005). Efficient Heuristics methods, such as the Monte Carlo Markov Chain (Hargreaves and Annan 2002), genetic algorithms (Price et al. 2006) and oracle based optimization (Beltran et al. 2006), require a lower number of experiments. However, these methods are sequential and if they were to be applied on a model such as FAMOUS, it would require 100 days to perform 100 experiments with one experiment a day. Annan and Hargreaves (2007) argues that the most efficient calibration method for climate models is the Ensemble Kalman filter which has been used on low complexity and low resolution climate models (Annan et al. 2005a, b). Applied to climate models, the Ensemble Kalman filter requires the use of an iterative scheme to increase the spread of the ensemble around its mean. Due to their architecture, modern cluster computers lend themselves to parallel rather than sequential or iterative schemes. Most climate models only scale to modest number of processors (in the case of FAMOUS it is 16–32 cpus) and more processors cannot efficiently increase the speed of the model. Our computational constraint is not the total CPU time but the wall-clock time it would take to perform the tuning. We therefore favoured the use of a parallel method rather than a sequential or iterative method. For that reason we chose to use a Latin hypercube sampling scheme to perform our tuning.

There are several issues with model calibration which should be taken into account. Complex models such as OAGCM output a large number of variables for which observational data is available (e.g. Temperature, precipitation, sea ice). The different diagnostics can be combined into one metric in different ways, but the choices of the metric used can influence the result of the tuning. As discussed in Rougier (2007), models are not a perfect representation of reality and uncertainty exists not only due to the lack of constraint on parameter values (parametric uncertainty) but also due to the nature of approximation that are made in the model (structural uncertainty). Parameter values can compensate for missing processes in the model and if we choose to include different variables in our metric then we can imagine a case where different combinations of parameter values result in the same optimal metric value. Therefore, performing a tuning which results in the selection of a single ‘standard parameterisation’ may not be ideal.

Using large ensembles to perform a single experiment can provide an estimate of the uncertainty in the result (Rougier 2007) but this requires a great amount of computational resources every time we want to perform an experiment. We choose to use a middle ground approach by which we do not restrict ourselves to selecting a single model configuration but choose to select a subset of experiments which represent ‘possibilities’.

We present here a tuning of FAMOUS (a low resolution version of the Hadley Centre climate model, HadCM3), performed using a Latin hypercube tuning method. We define a comprehensive cost function which takes into account seasonality and incorporates variables representing different aspects of the climate system. As well as using present day observations in our cost function, we choose to include a proxy for the climate of the LGM, in part because our model is particularly useful for paleoclimate studies and also because the LGM has been shown to provide additional constraint for the climate sensitivity of the models (Edwards et al. 2007; Annan et al. 2005c). After briefly describing the model in Sect. 2, we detail the tuning method in Sect. 3. Section 4 investigates how the definition of the cost function impacts on the results of the tuning and Sect. 5 describe the present day and LGM climates of the subset of selected runs.

2 Model description

FAMOUS is a low resolution ocean atmosphere GCM derived from the Hadley Centre coupled model HadCM3 (Gordon et al. 2000). Its resolution is roughly half of HadCM3’s, both in the ocean and the atmosphere, which makes it ten times faster to run than its parent HadCM3. The atmospheric resolution is 7.5° longitude by 5° latitude with 11 vertical levels and a time step of 1 h. The ocean resolution is 3.75° longitude by 2.5° latitude with 20 vertical levels and a time step of 12 h. Land processes are modelled with the Met. Office’s land surface scheme (MOSES1, Cox et al. 1999).

FAMOUS has previously been tuned in a systematic way (Jones et al. 2005) towards HadCM3 and then manually (Smith et al. 2008) to reduce the original northern high latitude cold bias. Our version of the model is different to the version of Smith et al. (2008): it uses a slightly different topography and uses two sweep time stepping (or double scan dynamics) to allow for a better numerical stability of the model under LGM boundary conditions. This different dynamical scheme introduces a significant warming in the northern high latitudes therefore the tuning of Smith et al. 2008 is not optimum for this version of the model.

For this tuning study we run sets of simulations with present day (PD) and LGM boundary conditions. The present day boundary conditions are identical to the version

of Smith et al. (2008) except for the orography which follows ICE-5G's present-day fields (Peltier 2004). For LGM boundary conditions, the orography and ice sheets extent are taken from ICE-5G reconstructions and the greenhouse gases and insolation values follow the PMIP2 standards (Braconnot et al. 2007).

3 Method

To perform our tuning, we chose to use a Latin hypercube sampling scheme which consists of (1) choosing the parameters to tune and defining the range of possible values, (2) sampling sets of parameter values within the parameter space and using these to perform an ensemble of experiment, (3) defining and applying a cost function which compares the output of experiments to observational data to determine optimum experiments (and associated parameter values).

3.1 The tuning parameters

3.1.1 Description

More than 100 parameters can potentially be tuned within a model like FAMOUS but varying a large number of parameters in a tuning increases the number of simulations to run. We have decided to tune ten parameters. These include the six parameters chosen by Jones et al. (2005) for the initial tuning of FAMOUS, that were chosen for having a high impact on the climate of HadCM3 (Murphy et al. 2004):

- RHcrit: the threshold of relative humidity for cloud formation (Smith 1990).
- Vf1: precipitating ice fall-out speed (Heymsfield 1977).
- Ct: the conversion rate of cloud liquid water droplets to precipitation (Smith 1990).
- Cw: the threshold values of cloud liquid water for formation of precipitation (Smith 1990). The value of this parameter is different for land and sea and will be varied together by the same fraction.
- Z0 (sea): the free convective roughness length over the sea for boundary layer processes (Smith 1993).
- Wave: gravity wave and trapped lee wave constants. These two parameters will also be varied together (Gregory et al. 1998).

Four parameters were added in this study:

- AlphaM: the sea ice low albedo (Crossley and Roberts 1995).
- Atm Diff: The horizontal atmospheric diffusion parameters varied together.

- Ocn H Diff: the oceanic horizontal diffusion parameters varied together.
- Ocn V Diff: the oceanic vertical diffusion parameters varied together.

The sea ice low albedo in FAMOUS decreases linearly with temperature within a specific range. Outside this range, the albedo is kept to a high value, for colder temperatures, and to low value (AlphaM), for warmer temperatures. This parameterisation accounts for the presence of melting ponds that form on the sea ice in summer. AlphaM was manually tuned in the study of Smith et al. (2008) and was set to a value of 0.2 which is lower than the range of values estimated in Murphy et al. (2004) for this parameter. We therefore included this parameter to our tuning.

The diffusion parameters were added to this tuning in order to improve the energy transport by the atmosphere and the ocean and thus reduce the cold northern bias present in the model as suggested by Jones et al. (2005).

We do not include the entrainment rate coefficient of the convection scheme in this tuning because this parameter is known to have a large impact on the structure of the atmosphere. As it is difficult to include any target for the vertical structure of the atmosphere, tuning this parameter could lead to a model with an unrealistic atmospheric structure.

The ranges of values for atmospheric and sea ice parameter were taken from Murphy et al. (2004), the ranges of the ocean and atmospheric diffusion coefficients were decided by performing stability tests, and the lower range of the sea ice low albedo is set to the value of Smith et al. (2008).

3.1.2 Preliminary sensitivity analysis study

We performed a preliminary study to determine the effect of individual parameters on the present day and LGM climates. We ran a set of single parameter perturbations where we change the value of each parameter to its maximum and minimum values. Our control simulation (CTRL) is our version of the model before tuning and has the same parameter values as Smith et al. (2008). The parameter values of CTRL correspond to intermediate values within the range of possible values except for the wave and ocean horizontal diffusion parameters which are set to their maximum possible value and for the sea ice low albedo which is set to the lower boundary of its range.

We perturbed each parameter to its maximum or minimum value taken from Murphy et al. (2004), and whenever the parameter values used in CTRL is already equal to the maximum or minimum value we also perturbed the parameter to an 'intermediate' value which corresponds to the

Table 1 Parameter perturbations and associated responses in global mean annual temperature for present day and LGM boundary conditions. The parameter change is calculated as the % change compared to the control values. The PD temperature anomaly is the

difference between annual average global mean temperature between a perturbed run and the control run for present day. The LGM sensitivity change is: $|T(\text{LGM}) - T(\text{PD})| - |T_{\text{ctrl}}(\text{LGM}) - T_{\text{ctrl}}(\text{PD})|$

Parameters		% parameter change	PD temp anomaly	LGM sensitivity change	Comments
RHCRT	min	−13	−1.68	−0.13	Increased values yield an increase of low clouds in the tropics and a decrease of higher clouds in medium and high latitudes. The result is a global warming higher over tropical land
	max	31	2.5	−0.28	
VF1	min	−71	3.18	−0.1	A decrease leads to an increase of high clouds. The greenhouse effect increases more than the planetary albedo leading to a global warming
	max	14	−0.11	−0.2	
CT	min	−47	−3.62	+0.45	A decrease reduces the total cloud coverage (especially high clouds) except over Indonesia. The planetary albedo increase leads to global cooling
	max	325			
CW	min	−43	−0.09	−0.19	An increase of CW leads to less precipitation and more low clouds. This leads to a higher planetary albedo which produces a global cooling
	max	1176	−4.9	+0.19	
ZOFSEA	min	−82	0.21	−0.22	High ZOFSEA values decrease the temperature over the ocean at medium and high latitudes
	max	350	−0.36	−0.26	
Waves	min	−50	0.19	+0.03	A minimum value causes changes in atmospheric circulation in northern high latitude leading to warmer conditions over Siberia and north Pacific and cooler conditions over north Atlantic in winter
	int	−25	0.06	−0.08	
AlphaM	int	150	−1.29	+0.69	A higher value produces a cooling over sea ice (particularly strong in the arctic) and an increase of sea ice especially in autumn
	max	225	−2.93	+1.08	
Atm. diff.	min	−17	−0.98	+0.01	
	int	−8	−0.65	−0.19	
Ocean vert. diff.	min	−80	−0.11	−0.02	
	max	10	−0.08	−0.24	
Ocean hor. diff.	min	−20	−0.09	+0.13	
	int	−10	−0.04	+0.01	

middle to the range (Table 1). These perturbed simulations were run for 200 years starting from a spun-up state of the control model. The change in parameter values and the response in PD and LGM temperatures are shown in Table 1.

The range of values varies quite a lot from one parameter to another, showing the difference in uncertainty and understanding in these parameters. The change in PD temperature does not necessarily reflect the magnitude of the parameter change, but rather the sensitivity of the model to it. Compared to other parameters, RHCRT has a small uncertainty in its parameter value, but strongly influences the global temperature for PD. On the other hand, CW was changed by an order of magnitude but only gives a temperature response twice as big as the perturbation of RHCRT.

We found that cloud and sea ice parameters influence temperature on a global scale changing the energy balance at the top of the atmosphere, whereas convection and gravity

wave affects temperature at regional levels. The sea ice low albedo is the parameter which has the largest effect on the difference in temperature between present day and LGM (i.e. LGM cooling): increasing the sea ice low albedo from 0.2 to 0.65 increases the LGM cooling by more than 1°C. This effect of the sea ice low albedo is investigated in more detail in Sect. 5.2. The effects of individual parameters on the present day climate are described more fully in Table 1.

3.2 Sampling strategy

To perform the tuning, we varied all the parameters simultaneously using a Latin hypercube Sampling, a stratified-random procedure which provides an efficient way of sampling variables (Mckay et al. 1979). With this sampling scheme, the number of samples should be at least ten times the number of parameters (Loeppky et al. 2009).

Table 2 Diagnostics used as a target in the tuning and datasets associated

Diagnostics	Unit	Target
Surface air temperature over land (temp)	°C	CRU CL 1.0 ^a
Precipitation rate over land (precip)	mm/day	CRU CL 1.0
Sea surface temperatures (SST)	°C	HadISST ^b
Sea ice concentrations	(0–1)	HadISST
Top of the atmosphere shortwave (SW)	W/m ²	ERBE ^c
Top of the atmosphere longwave (LW)	W/m ²	ERBE
Net heat flux at the surface of the ocean	W/m ²	NOC1.1a ^d
Wind stress at the surface of the ocean (TauU, TauV)	N/m ²	NOC1.1a

^a New et al. 1999, ^bUK Meteorological Office 2006, ^cHarrison et al. 1990, ^dGrist and Josey 2003

Since we vary ten parameters, we should sample at least 100 sets of parameter values. Uncertainty analysis studies performed on intermediate complexity models have used a sampling size an order of magnitude greater (Schneider von Deimling et al. 2006; Edwards and Marsh 2005) but performing an uncertainty analysis is beyond the scope of this study. Our aim is to find optimum sets of parameter values. Increasing the number of samples could improve the accuracy of the tuning but we are limited in the number of simulations we can perform, because the cost of running a GCM is high relative to the available computing resources. We therefore chose to sample 100 sets of parameters. We then use these sets of parameters to perform 100 simulation for present day and 100 simulations for the LGM. The ensemble of runs performed used more than 5,200 h of CPU time and produced 2 TB of raw data. This represent a substantial achievement with a model as complex as FAMOUS.

We chose to sample the parameter values uniformly over the parameter space (i.e. we define our prior as a uniform distribution of the parameter values within their range). This choice was made because previous tuning studies for FAMOUS used simple techniques that did not span the whole parameter space we therefore expected that parameter values far from those of the control simulations could minimise the cost function. The range of each parameter is thus divided into 100 equiprobable intervals (equally spaced in this case because we assume uniform distribution) and in each interval a value is randomly selected. The 100 values obtained for each parameter are randomly grouped with the values of the other parameters producing a total of 100 sets of parameter values.

Unlike the iterative method proposed by Jones et al. (2005) in the initial tuning of FAMOUS, our method enables us to cover the whole parameter space and to take into account the interdependency of the parameters by varying them all simultaneously.

Using the sets of parameters created we ran an ensemble of 100 FAMOUS runs with modern boundary conditions and 100 runs with glacial boundary conditions. All simulations started from the spun-up control model conditions

and were run for 200 years. Mean climatologies are computed over the last 30 years of the runs. A cost function was then applied to calculate their ranking and a subset of 13 simulations were extended to 1,000 years to bring them to equilibrium.

3.3 The definition of the cost function

3.3.1 The target of the tuning

We chose to compare our model to climatological datasets. As in Jones et al. (2005), we include a wide range of diagnostics to avoid the risk of improving one aspect of the model output at the expense of another. For present day, our diagnostics include well known climatic parameters such as temperature and precipitation rate but also diagnostics relating to the energy balance of the model both at the top of the atmosphere and at the surface of the ocean. The model diagnostics we chose are stated in Table 2 along with the source dataset used. Where possible, each dataset was chosen carefully to avoid introducing artificial constraints such as reanalysis data. For example, in areas poorly covered by observations, using reanalysis data would potentially result in tuning our model towards another model. Some of the climatologies used here are poorly constrained in some regions of the globe. This is the case for the National Oceanography Center NOC 1.1 climatology in the southern ocean. We will show in the next section how we will deal with such uncertainties by adapting the weights in the cost function.

Assessing the ability of a model to simulate glacial climate is a more difficult task than for the present day climate. Very little data is available for this period and the uncertainties associated with climate reconstructions from proxies are large and difficult to evaluate. We therefore concentrate on the sea surface temperatures (SST) which have been carefully reconstructed within several international projects. We use annual SST anomalies from the Multiproxy Approach for the Reconstruction of the Glacial Ocean surface reconstruction (MARGO project members 2009) which provides a global reconstruction of the SSTs

using different proxies with an indication of uncertainty. The uncertainty associated with this reconstruction is particularly important in the North Atlantic basin where there is a large discrepancy between the temperatures reconstructed from different proxies. The Southern Ocean basin has a poor coverage in terms of annual mean temperature reconstruction. We therefore only consider temperature reconstruction on the tropical region between 40°N and 40°S.

3.3.2 The metric

As a large number of simulations are performed, it is necessary to define a metric or cost function to evaluate the difference between model output and observations in a single number. We chose to use a weighted version of the ‘Arcsin Mielke’ score (AMS; Watterson 1996) that was chosen by Jones et al. (2005) in the initial tuning of FAMOUS. It takes into account several aspects of the field to be compared. It is expressed as follows:

$$\text{AMS} = \frac{2}{\pi} \arcsin \left(\frac{2\rho}{\sigma + 1/\sigma + b^2} \right) \quad (1)$$

b is the normalised bias between the two fields given by

$$b = \frac{\bar{x} - \bar{y}}{\sqrt{s_x s_y}} \quad (2)$$

where x and y represent the latitude-longitude field of observations and model output for the same variable, \bar{x} is the area weighted mean and s_x is the spatial standard deviation. σ is the ratio of the spatial standard deviations s_x and s_y . ρ is the pattern correlation coefficient defined by

$$\rho = \frac{\bar{x'y'}}{s_x s_y} \quad (3)$$

Values are normalized from -1 to $+1$ where $+1$ is obtained for a perfect agreement between the two fields and -1 for anticorrelated fields. Values close to 0 or below indicate bad agreement. To use this score, both of the fields must be defined on the same grid points. Thus, the observational fields from the datasets are regridded onto the model grid and a mask is applied so that both fields have the same spatial distribution.

The climate resulting from the previous tuning of FAMOUS had a strong bias in the northern high latitudes which was especially important in winter. The cost function has thus been adapted in an attempt to reduce this bias. The score is calculated for each month and averaged over the year in order to take into account the seasonal cycle. To emphasise high latitudes, three regions are defined as followed: southern high latitudes (90°S to 40°S), tropics (40°S to 40°N) and northern high latitudes (40°N to 90°N).

Table 3 Weights used in the cost function from the different diagnostics and regions, expressed as a percentage of the total score

	North	Tropical	South	Total
Temperatures	13.3	11.0	0.8	25.2
Precipitations	8.0	6.6	0.5	15.1
SST	4.1	6.2	3.3	13.6
Sea ice	8.9	0.0	5.9	14.8
SW	2.2	5.0	2.0	9.2
LW	2.5	4.8	2.0	9.3
Surface heat flux	0.5	1.7	0.4	2.6
TauU	0.2	0.8	0.2	1.2
TauV	0.2	0.8	0.2	1.2
LGM SST	0.0	7.9	0.0	7.9

The top plot shows the scores for all the acceptable simulations and the bottom plot zooms in the higher scores to show the spread of the points

We defined our cost function as the average of the score for each month, each region and each diagnostics using weights specified in Table 3. We determined the weights taking into account different criteria: (1) the importance of the regions by applying a coefficient of 1.5 for the North, 1 for the Tropics and 1 for the south, (2) the importance of each diagnostic by putting more emphasis on temperatures, precipitation, SSTs and sea ice concentrations, (3) the relative number of grid cells covered by data in each region, and finally (4) the reliability of the data in each region which is interpreted from the literature into a coefficient. These weights are subjective and can be adapted to specific needs without the need to rerun simulations which is an advantage of this technique. We are specifically interested in tuning FAMOUS to well represent the high latitudes so that it can be coupled successfully to an ice sheet model, we therefore applied higher weights to the north region than to the tropics. We also apply a low coefficient in the southern region to the wind stress at the surface of the ocean (TauU and TauV) to reflect the higher uncertainty in the dataset. The sensitivity of the results to the weights used is discussed in Sect. 4.3.

4 Investigating the ensemble of simulations

4.1 An overview of the ensemble of experiments

Before applying our cost function to identify simulations that agree best with the observations, it is important to understand how it behaves when we apply it to our targets. Figure 1 shows the AMS scores obtained from the ensemble of models for each diagnostic. As noted by Watterson (1996), fields with a strong north–south gradient, such as surface air temperature over land and SST, generally have a high score,

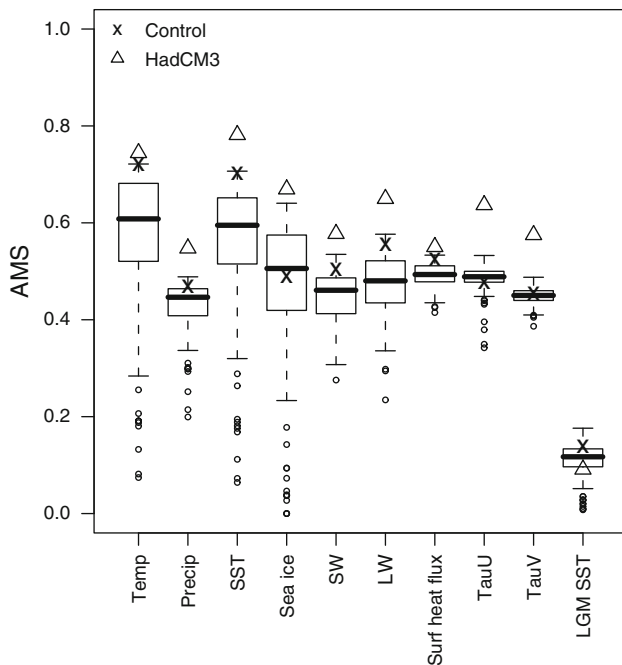


Fig. 1 Box-and-whisker plot showing the distribution of AMS scores for all the diagnostics. The *black crosses* represent the scores of the control model and the *triangles* represent the scores obtained for HadCM3 regridded onto the FAMOUS grid for comparison

whereas scores for precipitation are much lower: any shift in precipitation pattern results in a lower score. Diagnostics related to the energy budget also have lower ranges of scores than temperatures and sea ice diagnostics. This should be taken into account when determining the rank of the simulations and weights can be chosen to compensate for this effect. In this study, we value temperature, precipitation, and sea ice more than energy fluxes at the ocean surface.

The control experiment is amongst the top scores for most of the diagnostics (Fig. 1) except for the sea ice, for which there are 53 simulations with better scores. For most of the diagnostics it is possible to obtain a higher score than the control simulation by choosing a different set of parameters. Surface air temperature is the only diagnostic for which the score of the control model is not surpassed but we find simulations with a similarly good temperature scores.

For comparison, we show the score obtained with HadCM3 calculated on the same grid as FAMOUS. The scores obtained by the lower resolution model are generally lower than the scores of HadCM3, except for the LGM SSTs where 78 members of the ensemble have higher LGM scores.

Some parameter combinations result in very unrealistic climates with global mean temperatures range from 5 to 38°C. Some of the simulations where the climate has been pushed far from the initial state of the control simulation are still drifting considerably after 200 years. Rather than continuing those runs, we have decided to exclude them since our goal is to find simulations which are as similar as

possible to observational data. We therefore only take into account models with a present day global mean temperatures of $14 \pm 5^\circ\text{C}$. A total of 73 out of 101 models fall into this category (henceforth referred to as ‘acceptable models’).

Using the weights of Sect. 3.3, we can calculate the total score, and rank the models from highest to lowest scores. The total scores for ‘acceptable’ models range from 0.43 to 0.57. The control simulation comes at rank 14 with a score of 0.55. There are therefore 13 simulations which have a higher score than the control simulations and are therefore considered ‘better’ according to our criteria. We choose to select the subset of top 13 runs on the basis that they have higher scores than the control simulation and we define this subset as the ‘better’ models.

4.2 No clear optimum in the parameter values

We have evaluated whether a region of optimum parameter values can be identified. Figure 2 shows the AMS scores against each of the parameter coefficients, normalised from 0 for the minimum value to 1 for the maximum value. At first sight, it seems that lower values of RHCRT and higher values of VF1 give higher scores. For the other parameters, however, no clear optimum can be found. Parameters seem to compensate for each other so that very different combinations of parameter values can give similar AMS scores. This emphasises the benefits of using an objective tuning method over the more common hand tuning method. Performing more simulations or using a climate emulator as in Rougier et al. (2009) and Murphy et al. (2007) would be necessary to make any further conclusions on the relationship between parameter value and the score of a model. Moreover, the nature of the cost function used could also have an impact on the relationship between parameter values and score. Our cost function consists of adding scores obtained for different diagnostics and we may have different optima for the different diagnostics (i.e. precipitation and sea ice). This could lead to having a lot of local optimums in the parameter space.

4.3 Sensitivity of the result to the definition of the cost function

In order to evaluate if the results of the tuning are dependent on the overall cost function, we applied different weights to our cost function and compared the scores obtained for each simulation. We computed a simple cost function where the diagnostics are weighted only by the number of grid points containing data. Figure 3 compares this simple cost function to the one described in Sect. 3.3. There seems to be a linear relation between the result of the two functions showing that they both identify ‘good’ and ‘bad’ simulations in a similar way. However, focusing on

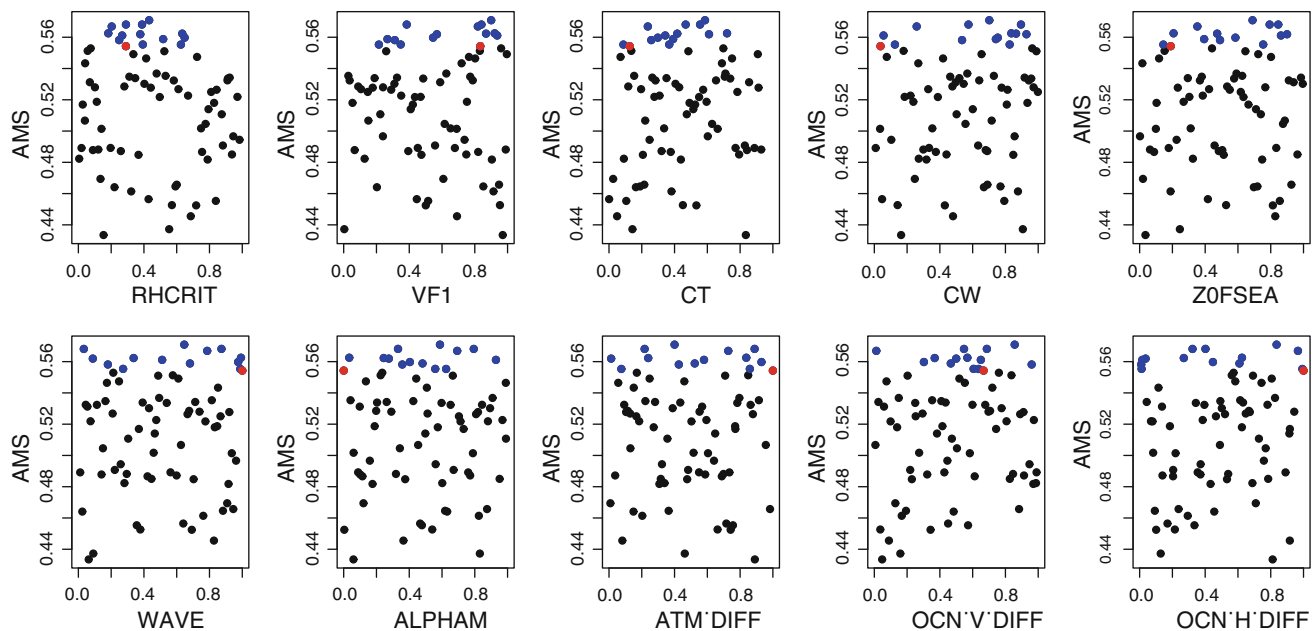


Fig. 2 Score of the ensemble plotted against parameter values. Parameter values are scaled from 0 for the minimum value to 1 for maximum value. Only acceptable models are represented here. The control model is shown in *red* and the top 13 models are shown in *blue*

just the very top scores, the points are more scattered. This means that using different weights changes the ranking of the simulations within the ensemble but without significantly changing the subset of top simulations (the subsets of top 14 simulations determined by the two scores only vary by 2 simulations). Therefore, the choice of cost function can influence the result of traditional methods of tuning which result in the selection of one model configuration but has less impact on our tuning method.

4.4 The effect of including LGM data in the cost function

Adding glacial constraints doubles the number of simulations to run for this tuning. It is thus important to verify if using glacial data adds an additional constraint on determining ‘good’ simulations. To test that, we compare the LGM score (anomalies of tropical SST), to the score for present day tropical SSTs (Fig. 4). Simulations with high PD tropical SST scores have a wide range of LGM scores and only a subset of them perform well during the LGM. Therefore, the LGM data clearly adds a further constraint on the tuning of the model and shows the benefits of using this broader range of tuning targets. We will investigate the benefits of using further palaeo targets in future work. Another advantage of performing those LGM simulations is that we can look at characteristics of the LGM climate such as the global temperature signal or the ocean circulations, as demonstrated in Sects. 5.2 and 5.3.

5 The subset of selected simulations

As noted in Sect. 4.1, there are 13 simulations which have a higher score than the control. Since 200 years is a relatively short time for coupled ocean atmosphere GCM simulations to get to equilibrium we extended the length of the top 13 simulations to a total of 1,000 years of integration. After 1,000 years of integration, the trend in surface air temperatures (calculated over the last 200 years of the runs) are small: in all of the present day and LGM simulations the trend are less than 0.12°C per century and in most cases less than 0.07°C per century. We therefore conclude that after 1,000 years of integration, the simulations are close to equilibrium. We then calculated the climatologies over the last 30 years and recalculated the cost function. In 4 of these 13 simulations, the climate continued to drift after the initial 200 years resulting in a lower overall cost function than the control. We therefore reject these four models and the new subset of the top nine simulations is now defined as the ‘good models’.

5.1 A great variety of behaviours amongst the ‘good’ simulations

Figure 5 shows the scores of the top 10 models (e.g. the ‘good’ models and the control model) obtained in each diagnostics compared with each other. These models have a great variety of performance. The control simulation has the strongest score for temperature diagnostics but quite a weak score for precipitation and energy budget compared

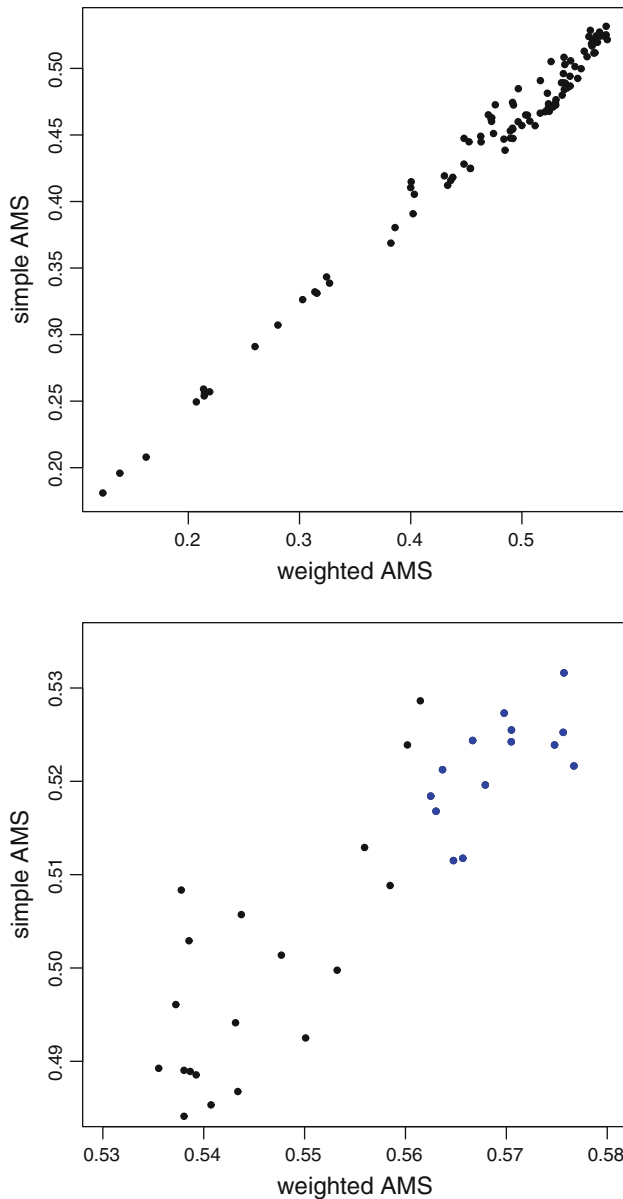


Fig. 3 Comparison of simple area weighted with the fully weighted cost function as described in Sect. 3.3

with the other ‘good’ models. The top simulation on the other hand, has a better score for sea ice and precipitation but a lower score for SSTs. Rather than obtaining a single simulation which has optimised the cost function for all the diagnostics, we have a variety of simulations with individual strengths and weaknesses while all having equally good overall scores.

Figure 6 shows a map of the range of present day annual mean temperature simulated within the subset of top ten models. The temperature variability within the subset is higher over continents and over area covered with sea ice.

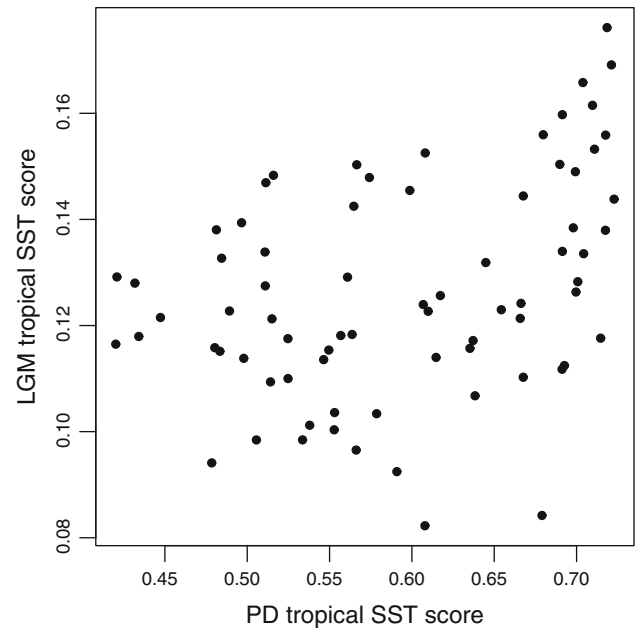


Fig. 4 Comparison of LGM and present day tropical SST AMS scores. The LGM score adds an extra constraint on the tuning

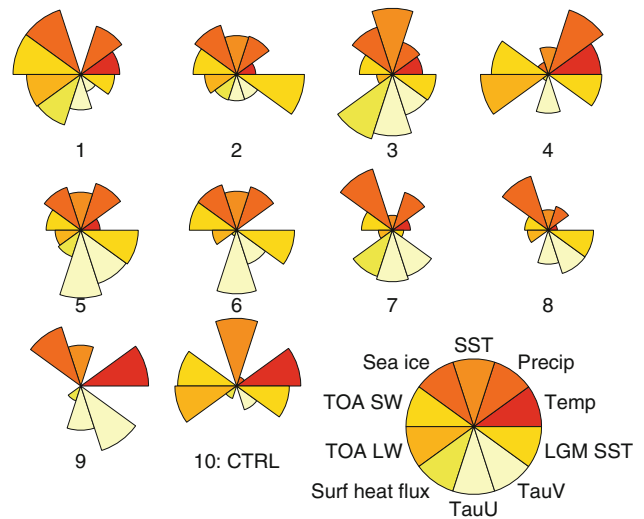


Fig. 5 Representation of the scores of the top ten models of the ensemble relative to each other. Each segment plot represents one model and the models are ranked according to their total score. The shades of red represent the weights applied to the diagnostics (see Table 3). The segments represent the score obtained by the models for each diagnostics scaled over the top ten models. A full radius and a null radius correspond, respectively, to the highest and the lowest score within the top ten runs. The control model identified as CTRL is the tenth best simulation. Simulations number 1 and 4 are identified in Sect. 5 as S1 and S4

In the Arctic region, the difference between individual models in the top ten is more than 5°C. This variability in the subset is especially high in winter.

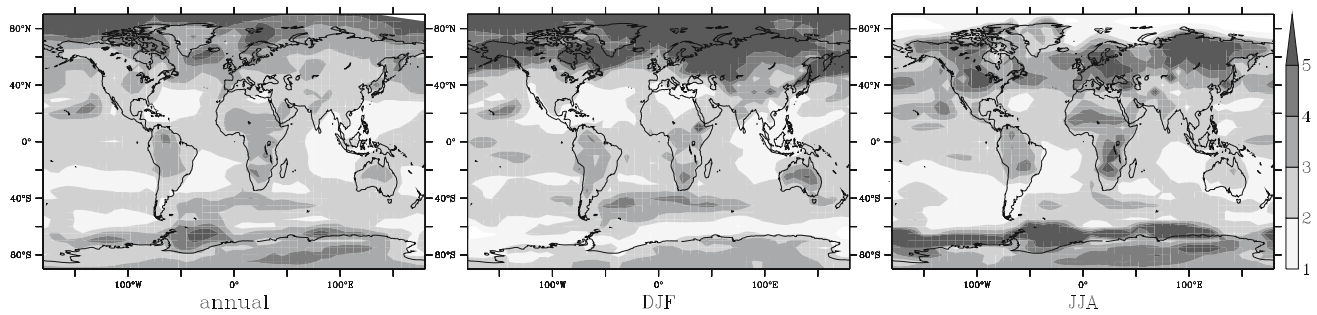
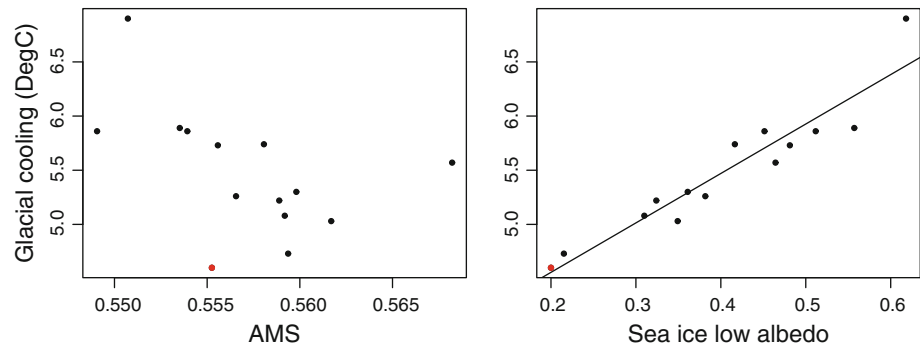


Fig. 6 The range of present day temperatures (in K) obtained within the top ten models for each grid point defined as $[\max(T) - \min(T)]$ for annual mean (annual), December to February (DJF) and June to August (JJA) means

Fig. 7 Glacial cooling of the ensemble of runs against, the scores of the simulations on the *left* and the sea ice low albedo on the *right*. The glacial cooling is defined as $|T(\text{LGM}) - T(\text{PD})|$



5.2 LGM temperature response within the selected runs

We define glacial climate sensitivity (or glacial cooling) as the absolute difference between present day and LGM annual mean surface air temperature. In other words, it is the global temperature response to the LGM forcing. The LGM cooling for the subset of top ten models is between 4.6 and 5.7°C, with lowest glacial cooling obtained by the control experiment. This result is within the range of the PMIP2 results (which is 3.6 to 5.7°C) amongst state-of-the-art ocean atmosphere coupled GCMs (Braconnot et al. 2007).

We tested whether the glacial cooling depends on any particular parameter. We found that the glacial cooling varies linearly with the sea ice low albedo (Fig. 7) with a correlation coefficient of 0.94. The impact of this sea ice parameter on glacial cooling was already highlighted in Table 1 which showed that a change of this parameter from its minimum to maximum value increased the glacial sensitivity (or glacial cooling) by 1°C.

Increasing the sea ice low albedo has the effect of increasing the amount of sea ice in summer but does not change the amount of sea ice in winter. This parameter only acts when the temperatures are warm. It therefore increases the reflectivity of the summer sea ice, which cools the atmosphere above, and results in more sea ice in summer. Because there is more sea ice in summer, there is also more sea ice in the subsequent autumn, even though the sea ice albedo in autumn is not affected by the change

of parameter value. At the LGM, the effect is even greater because the sea ice cover is greater. As a result, the cooling at the LGM is greater than the cooling for the present day which explains the link between glacial cooling and the sea ice low albedo.

5.3 The glacial ocean circulation of the selected runs

We investigate the Atlantic Meridional Overturning Circulations (AMOC) in the ensemble, under PD and LGM boundary conditions. Among the ‘good’ models, the present day maximum strength of the AMOC varies from 15.8 to 18.8 Sv, which is within to the range of observational estimates of 18 ± 3 –5 Sv (Talley et al. 2003).

To evaluate the response in AMOC under glacial boundary conditions, we determine for each run the changes of depth of the North Atlantic Deep Water (NADW) (which is calculated as the change of depth of the 0 Sv contour of the AMOC at the equator) and the changes in the maximum of the stream function between present day and LGM runs. Figure 8 shows the values obtained for the top ten models. With the exception of one model, ‘good’ models all show a weakening of the AMOC with generally slightly shallower NADW. The control model has a strengthening and deepening of the AMOC and a very weak Antarctic bottom water cell. Palaeo-proxies suggest that the AMOC was of comparable strength or slightly slower than today with a NADW cell shallower than today (McManus et al. 2004; Lynch-Stieglitz et al.

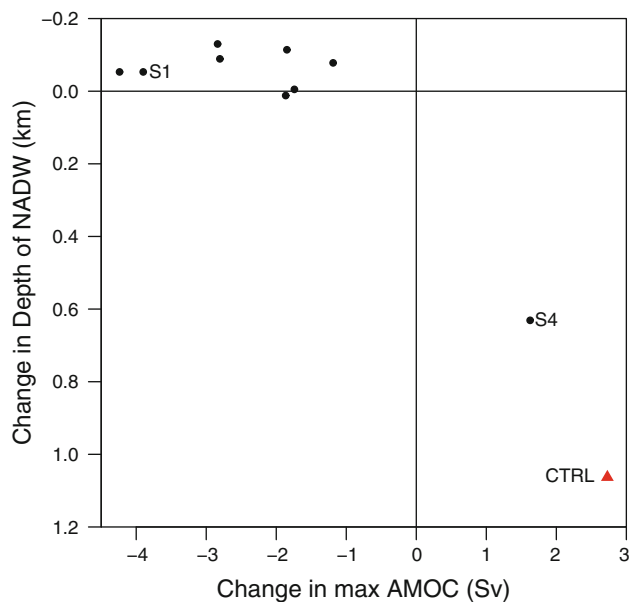


Fig. 8 Changes in AMOC between present day and LGM (LGM–PD) for the top ten simulations including CTRL represented as a *red triangle*. The depth of the NADW is defined as the depth of the 0 contour of the AMOC at the equator

2007). This behaviour of the glacial NADW is followed by all but one of the nine selected models which is analysed in the next section. Our ensemble of ‘good’ models is therefore more in agreement with proxy data than the control simulation.

5.4 Improving the simulation of present day precipitations

As shown in Fig. 5, the control simulation has the lowest precipitation score in the top ten simulations. In this section, we have a closer look at simulation number 4 (S4), which obtained the highest score in precipitation and outgoing longwave at the top of the atmosphere, to understand the link between parameter values, clouds and score.

S4 has relatively similar parameter values to the control experiment except for greatly enhanced CT and CW (Fig. 9). This simulation has slight changes in other cloud parameters, such as RHcrit and VF1, but most importantly it has the same value for the sea ice low albedo.

The climate obtained in S4 is colder than the control simulation over the mid and high latitudes northern hemisphere continents and over the sea ice. This cooling over northern hemisphere continents happens in summer (see Fig. 10) and autumn. Because the summers are cooler, there is an increase in sea ice in autumn which produces a cooling over the Arctic sea ice during autumn. The summer cooling over northern hemisphere land is due to an increase in the amount of low clouds (Fig. 11) which provides

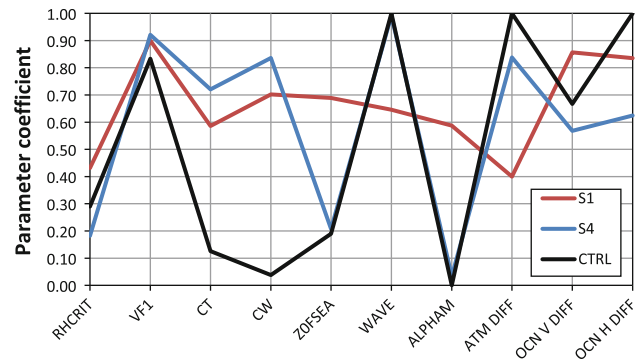


Fig. 9 Parallel plot showing the parameter values for CTRL, S1 and S4 models. The parameter values are scaled from 0 to 1, 0 corresponding to the minimum value and 1 corresponding to the maximum value

additional shading without increasing the greenhouse effect.

It is the combination of the increase in CT and CW which produces this increase in low clouds over land. These two parameters are both used in the equation that determines the amount of precipitation in clouds from their amount of liquid water. CT determines the rate at which water precipitates but only when the cloud liquid water content is high compared to CW. So the two parameters act in opposite direction. The change in cloud happens over land because the value of CW is higher over land than over sea to account for the difference in the size of droplets. In our tuning we varied the land and sea values of CW together by the same coefficient. As a result CW over land is increased much more than over sea. CW and CT compensate each other over sea, but the effect of CW is greater than the effect of CT over land, leading to a reduction in the precipitation rate over land only. The change in climate happens in summer because during this season, the relative humidity is lower, therefore the cloud liquid water content is lower and closer to the threshold controlled by CW. Summer conditions thus maximise the effect of CW.

Figure 10 shows the difference between S4 and observation and between the control and observation for the temperature and precipitation in summer (June, July and August average). We can see that the errors in temperature are not reduced compared to CTRL which is consistent with the score for temperature obtained. We go from a warm bias in the northern high latitude continents in CTRL to a cold bias. The errors in precipitation on the other hand seem to be reduced: there is less excess of precipitation on northern high latitude continents and the errors in the ITCZ are reduced especially in the West Pacific due to the shift in the ITCZ and the increase in precipitation over the West Pacific.

S4 and CTRL have very similar responses to LGM boundary conditions. The glacial cooling of S4 is similar to

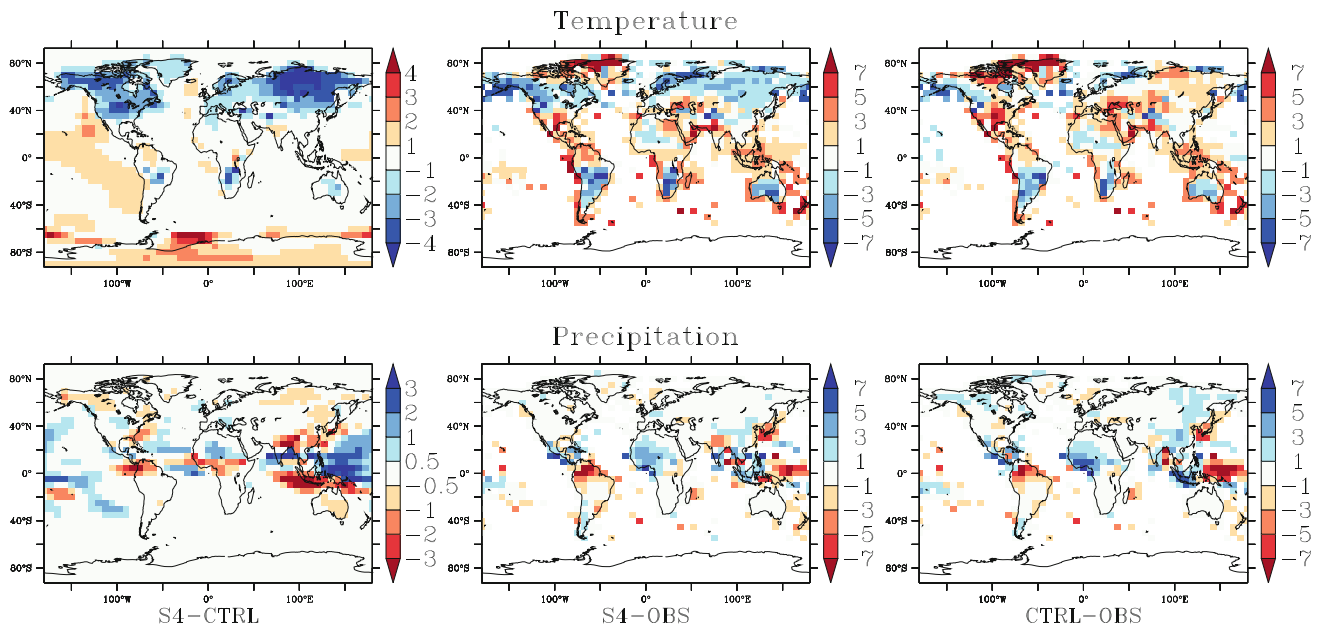


Fig. 10 Difference in summer (June, July and August mean) temperature and precipitation between S4, CTRL and CRU climatology data (New et al. 1999)

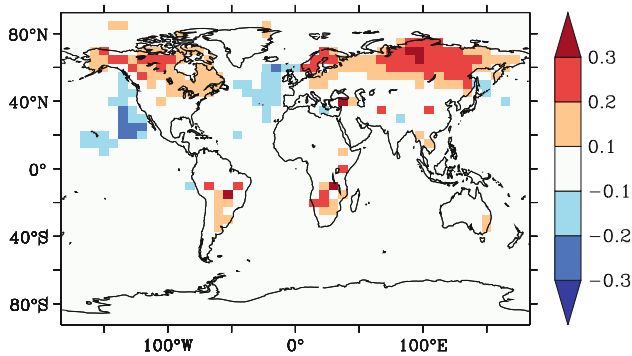


Fig. 11 Difference in summer (June, July and August mean) low cloud cover in grid cells between S4 and CTRL

the CTRL because they have very similar values for the sea ice low albedo. As a result, their LGM sea ice extent is comparable, and in both simulations, there is an increase in the glacial Atlantic overturning circulation compared to the present day AMOC. Therefore, the parameter combination in S4 modifies the present day climate but does not change its sensitivity to glacial boundary conditions.

5.5 Improving present day sea ice and the effect on LGM climate

The experiment which has the best sea ice score ranks the highest (S1). The precipitation and longwave fields are slightly improved compared to the control and the net surface heat flux is greatly improved (see Fig. 5). Temperature and SSTs scores on the other hand are slightly

lower than CTRL. This is the simulation which, according to our criteria, is the most balanced. Most of the parameters in this simulation are different to those of the control simulation, in particular the sea ice low albedo is increased (Fig. 9).

As in the simulation S4, we observe a cooling over land in summer above 40°N compared to CTRL (Fig. 12). This cooling is not as high as in S4 but since the pattern and season correspond, it could be due to the effect of the combined increase in CT and CW as described in previous section. We also observe a cooling over sea ice in summer and autumn (Fig. 12). This cooling is related to an increase in sea ice cover in summer and autumn in the northern hemisphere except in the Nordic sea (Fig. 13). This can be attributed to the effect of the sea ice low albedo as explained in Sect. 5.2. We observe seasonal shifts in the ITCZ linked with the seasonal changes in temperatures and a general increase in the precipitation in the tropics (Fig. 12). As for S4, the errors in the precipitation field are reduced but the errors in the temperature are increased. In particular the tropics in S1 are 1–2°C warmer than observations (Fig. 12).

At the LGM, S1 has a lower tropical SST score than the control simulation but its glacial AMOC is more in agreement with proxies as it is slower and slightly shallower than at present day. The maximum overturning is reduced from 18 Sv at present day to 14 Sv for the LGM, and the sea ice cover is increased compared to the control LGM simulation. Sea ice area is larger all year long but is especially increased in late summer and early autumn (see

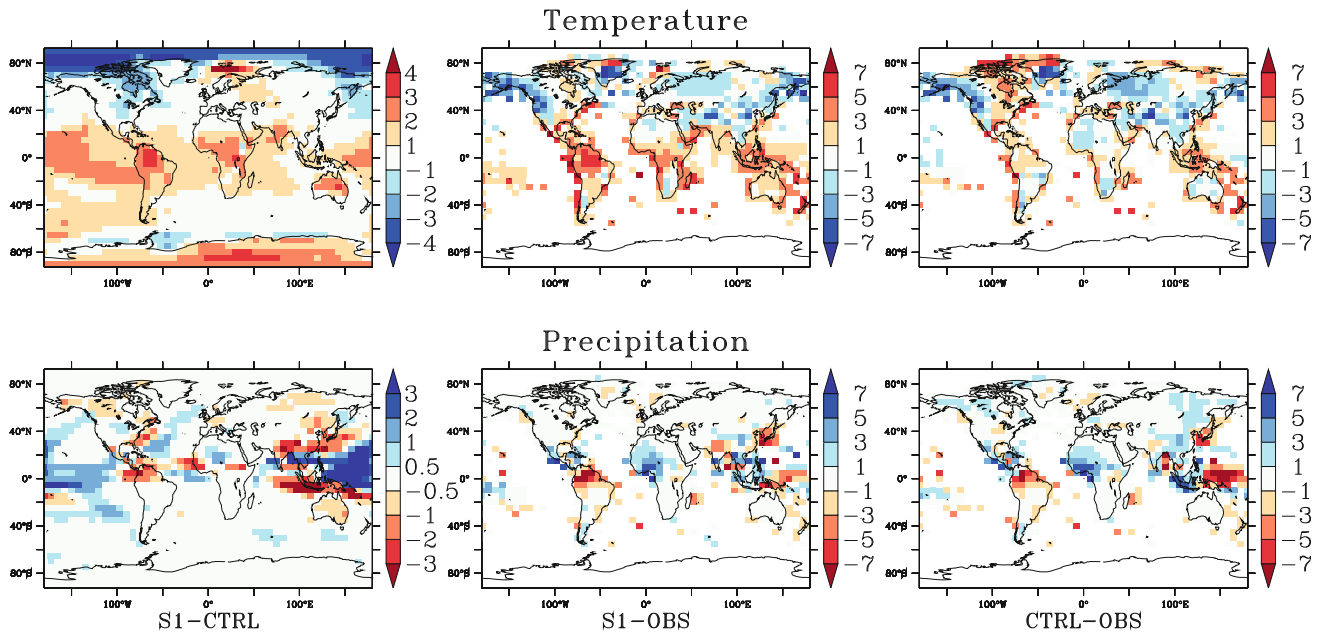


Fig. 12 Difference in annual mean temperature and precipitation between S1, CTRL and CRU climatology data (New et al. 1999)

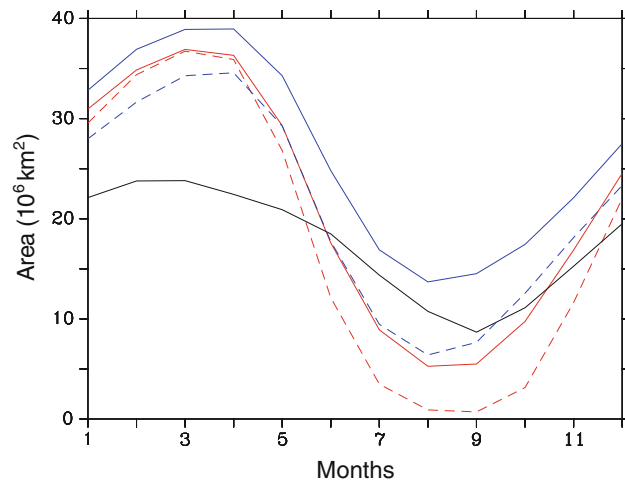


Fig. 13 Annual cycle of the arctic sea ice cover at present day (in red) and LGM (in blue) in S1 (solid lines) and CTRL (dashed lines). The black line corresponds to the present day observations from HadISST (UK Meteorological Office 2006)

Fig. 13). The deep water formation in the north Atlantic occurs further south than in the LGM control, due to the increase in the sea ice extent. Finally, the glacial cooling is increased compared to CTRL, because the sea ice extends further at the LGM than at present day.

In the S1 simulation, the combination of parameter values result in an improvement of the overall present day climate according to our metric. This combination of parameter values also substantially affects the LGM climate and in particular improves the overturning circulation. As we showed here, this change in LGM climate is

linked to the effect of the sea ice and is likely caused by the change in sea ice albedo.

6 Conclusion

We have tuned a low resolution GCM using Latin hypercube sampling. This method enables us to investigate the whole parameter space by taking into account the interdependencies between the parameters. The method is easy to implement and offers great flexibility by allowing all model experiments to be run in parallel. It is therefore well adapted to the use of modern computer clusters.

The ranking of the models are then determined by a cost function which compares the model output to present day and LGM data. This cost function can easily be adapted to specific needs by putting more emphasis on some diagnostics, and taking into account the uncertainty in the dataset used. In theory, different cost functions can be used on the same ensemble (without the need for additional experiments), optimising the use of the model for different purposes.

The ‘objective’ tuning method we present, along with other parameter estimation techniques, encompasses a bigger range of tuning options than the traditional hand tuning. It still necessitates subjective choices which are driven by ‘expert solicitation’, such as the choice of parameters selected for tuning, the range of values spanned by the sampling and the definition of the cost function. But these choices are made clear during the process, and the definition of the tuning problem is greatly improved.

Although including glacial data in our cost function necessitates running the ensemble of models with glacial boundary conditions, we show that it offers additional constraints on the tuning. Implicit within using the LGM as a tuning target is that we are tuning a model to the given set of forcings, and the tuning may also be compensating for some missing forcing (e.g. higher atmospheric dust concentrations).

We select a subset of top nine models defined as ‘good’ models which display a great variety of behaviours, but have a higher score than the standard control version of the model. Although the cost function applied is subjective, we show that weighting the target diagnostics differently does not greatly change the subset of ‘good’ runs obtained, but the ranking of the simulations differ. This effect of the choice of cost function on the ranking of simulations would influence the result of traditional methods of tuning where only a single solution is selected.

We investigated how the glacial sensitivity and the Atlantic overturning circulation vary within our ensemble of models. The control model has the lowest glacial sensitivity of the ensemble, due to a sea ice parameter which was tuned to improve present day climate. The ‘good’ runs display present day AMOC strengths that lie within the range of observational estimates. And most of the ‘good’ models have shallower and weaker glacial NADW than the control model, which is in better agreement with estimates from palaeo-proxies.

Most tuning exercises focus on improving the present day climate. We showed that including other climate regimes as targets such as the LGM leads to a different choice of tuned models. Using present day constraints is a necessary but not sufficient condition for accurate representation of past and future climates. Moreover, a single model cannot give a perfect representation of the climate due to the intrinsic structural uncertainty of GCMs. It is therefore necessary to consider more than one configuration. Our study provides a compromise between the use of big ensemble of models to investigate uncertainty in modelling (Stainforth et al. 2005; Murphy et al. 2004), and the constraints associated with the use of computationally intensive models such as state-of-the-art GCMs. Such subset of models have some benefits over the purely statistically based approaches in that the smallness of the subset allows investigations of the different possible responses of the climate to specific forcings as well as giving insight on the mechanisms operating. Examples of applications of our small ensemble of tuned models are freshwater hosing experiments, to investigate the range of response of the climate model to freshwater forcing under present day or LGM boundary conditions, and ice sheet forcing, to analyse the sensitivity of the Northern Hemisphere LGM ice sheets to climate forcing.

Acknowledgments We thank our reviewers for their useful comments. This work was funded by the European Marie Curie Research Training Network for Ice sheet and Climate Evolution (NICE) and was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol, <http://www.bris.ac.uk/acrc/>.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Annan J, Hargreaves J (2007) Efficient estimation and ensemble generation in climate modelling. *Philos Trans Royal Soc A Math Phys Eng Sci* 365(1857):2077–2088
- Annan JD, Hargreaves JC, Edwards NR, Marsh R (2005a) Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean Model* 8(1–2):135–154
- Annan JD, Lunt DJ, Hargreaves JC, Valdes PJ (2005b) Parameter estimation in an atmospheric GCM using the Ensemble Kalman Filter. *Nonlinear Process Geophys* 12(3):363–371
- Annan JD, Hargreaves JC, Ohgaito R, Abe-Ouchi A, Emori S (2005c) Efficiently constraining climate sensitivity with ensembles of paleoclimate simulations. *SOLA* 1:181–184
- Beltran C, Edwards N, Haurie A, Vial J, Zachary D (2006) Oracle-based optimization applied to climate model calibration. *Environ Model Assess* 11(1):31–43
- Braconnot P, Otto-Bliesner B, Harrison S et al (2007) Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum. Part 1. Experiments and large-scale features. *Clim Past* 3(2):261–277
- Cox PM, Betts RA, Bunton CB et al (1999) The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Clim Dyn* 15(3):183–203
- Crossley JF, Roberts DL (1995) Thermodynamic/dynamic Sea-ice model. Meteorological Office, Bracknell
- Edwards NR, Marsh R (2005) Uncertainties due to transport-parameter sensitivity in an efficient 3-D ocean-climate model. *Clim Dyn* 24(4):415–433
- Edwards TL, Crucifix M, Harrison SP (2007) Using the past to constrain the future: how the palaeorecord can improve estimates of global warming. *Prog Phys Geogr* 31(5):481–500
- Gordon C, Cooper C, Senior CA et al (2000) The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim Dyn* 16(2):147–168
- Gregory D, Shutts GJ, Mitchell JR (1998) A new gravity-wave-drag scheme incorporating anisotropic orography and low-level wave breaking: impact upon the climate of the UK Meteorological Office Unified Model. *Q J Royal Meteorol Soc* 124(546):463–493
- Grist JP, Josey SA (2003) Inverse analysis adjustment of the SOC air-sea flux climatology using ocean heat transport constraints. *J Clim* 16(20):3274–3295
- Hargreaves JC, Annan JD (2002) Assimilation of paleo-data in a simple Earth system model. *Clim Dyn* 19(5–6):371–381
- Harrison EF, Minnis P, Barkstrom BR et al (1990) Seasonal-variation of cloud radiative forcing derived from the earth radiation budget experiment. *J Geophys Res Atmos* 95(D11):18687–18703
- Heymsfield AJ (1977) Precipitation development in stratiform ice clouds—microphysical and dynamical study. *J Atmos Sci* 34(2):367–381

- Jones C, Gregory J, Thorpe R et al (2005) Systematic optimisation and climate simulation of FAMOUS, a fast version of HadCM3. *Clim Dyn* 25(2–3):189–204
- Loeppky JL, Sacks J, Welch WJ (2009) Choosing the sample size of a computer experiment: a practical guide. *Technometrics* 51(4):366–376
- Lynch-Stieglitz J, Adkins JF, Curry WB et al (2007) Atlantic meridional overturning circulation during the last glacial maximum. *Science* 316(5821):66–69
- MARGO project members (2009) Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum. *Nat Geosci* 2(2):127–132
- Mckay MD (1992) Latin hypercube sampling as a tool in uncertainty analysis of computer models. In: *Proceedings of the 24th conference on winter simulation*. ACM Press, Arlington, Virginia, pp 557–564
- Mckay MD, Beckman RJ, Conover WJ (1979) Comparison of 3 methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21(2):239–245
- McManus JF, Francois R, Gherardi JM, Keigwin LD, Brown-Leger S (2004) Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature* 428(6985):834–837
- Murphy JM, Sexton DMH, Barnett DN et al (2004) Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* 430(7001):768–772
- Murphy JM, Booth BB, Collins M et al (2007) A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Philos Trans Royal Soc A Math Phys Eng Sci* 365(1857):1993–2028
- New M, Hulme M, Jones P (1999) Representing twentieth century space-time climate variability. Part 1. Development of a 1961–1990 mean monthly terrestrial climatology. *J Clim*, pp 829–856
- Peltier WR (2004) Global glacial isostasy and the surface of the ice-age earth: the ice-5G (VM2) model and grace. *Annu Rev Earth Planet Sci* 32:111–149
- Price AR, Voutchkov II, Pound GE et al (2006) Multiobjective tuning of grid-enabled earth system models using a non-dominated sorting genetic algorithm (NSGA-II). In: *Proceedings of the 2nd international conference on escience and grid computing*. Amsterdam
- Rougier J (2007) Probabilistic inference for future climate using an ensemble of climate model evaluations. *Clim Change* 81(3):247–264
- Rougier J, Sexton DMH, Murphy JM, Stainforth D (2009) Analyzing the climate sensitivity of the HadSM3 climate model using ensembles from different but related experiments. *J Clim* 22(13):3540–3557
- Schneider von Deimling T, Held H, Ganopolski A, Rahmstorf S (2006) Climate sensitivity estimated from ensemble simulations of glacial climate. *Clim Dyn* 27(2):149–163
- Smith RNB (1990) A scheme for predicting layer clouds and their water-content in a general-circulation model. *Q J Royal Meteorol Soc* 116(492):435–460
- Smith RNB (1993) Subsurface, surface and boundary layer processes. Meteorological Office, Bracknell
- Smith RS, Gregory JM, Osprey A (2008) A description of the FAMOUS (version XDBUA) climate model and control run. *Geosci Model Dev* 1(1):53–68
- Stainforth DA, Aina T, Christensen C et al (2005) Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* 433(7024):403–406
- Talley LD, Reid JL, Robbins PE (2003) Data-Based Meridional Overturning Streamfunctions for the Global Ocean. *J Clim* 16(19):3213–3226
- UK Meteorological Office H C (2006) HadISST 1.1—Global sea-ice coverage and SST (1870–Present). British Atmospheric Data Centre. Available from: <http://badc.nerc.ac.uk/data/hadisst/>
- Watterson IG (1996) Non-dimensional measures of climate model performance. *Int J Climatol* 16(4):379–391